Benchmarking Neuro-Symbolic Reasoners: Existing Challenges and A Way Forward

Gunjan Singh [a,*], Riccardo Tommasini [b], Sumit Bhatia [c] and Raghava Mutharaju [a]
[a] *Knowledgeable Computing and Reasoning Lab, IIIT-Delhi, Delhi, India*
E-mails: gunjans@iiitd.ac.in, raghava.mutharaju@iiiitd.ac.in
[b] *LIRIS Lab, INSA, Lyon, France*
E-mail: riccardo.tommasini@insa-lyon.fr
[c] *Media and Data Science Research Lab, Adobe Inc., Delhi, India*
E-mail: sumit.bhatia@adobe.com

Abstract. Neuro-Symbolic approaches bring together symbolic logic and neural network-based machine learning. This has the potential to build robust reasoning systems. However, the field faces challenges due to diverse design approaches and evaluation methods. We address the latter challenge by emphasizing the critical requirement for a comprehensive benchmark framework tailored to the unique evaluation needs of neuro-symbolic reasoning systems. We highlight the importance of such benchmarks and discuss essential criteria, including metrics, dataset selection, and the formulation of reasoning tasks. This work contributes towards a more systematic and principled evaluation framework for neuro-symbolic reasoning, highlighting the broader role of benchmarks in advancing the field.

Keywords: Benchmark, Neuro-Symbolic AI, Reasoning, Description Logics, Ontology, Neural Network

1. Introduction

Neuro-symbolic Artificial Intelligence (AI) [1, 2] is a promising field that aims to bridge the gap between traditional symbolic logic and modern neural network-based machine learning. The idea is to combine the strengths of both approaches while overcoming their weaknesses. The focus of this paper lies within the realm of neuro-symbolic reasoning. At its core, neuro-symbolic reasoning involves integrating symbolic reasoning, which relies on structured logic and formal knowledge representation, with neural network-based methods known for their capacity to process large-scale, unstructured data and learn complex patterns from it. This fusion holds the potential for developing systems with enhanced performance, explainability, and generalization abilities [3]. It's important to note that these approaches, unlike traditional reasoning methods, are not necessarily sound and complete. Instead, they strike a balance between approximating the precise reasoning capabilities of symbolic systems and harnessing the robust learning capabilities of machine learning techniques.

However, progress in this field faces significant challenges because neuro-symbolic reasoning is emerging, in contrast to other areas with extensive research and well-established benchmarks. For instance, several models (Graph Neural Networks (GNN) [4], Logic Tensor Networks [5]), methodologies (Inductive Logic Programming [6]) and innovative ideas (explainable AI [7], zero-shot learning [8]) enrich this field. As a result, existing works in this field exhibit diversity in techniques, and hence, different methods and criteria are used to evaluate the performance of

*Corresponding author. E-mail: gunjans@iiitd.ac.in.

neuro-symbolic reasoning systems (see Table 1). The lack of a standardized approach makes it difficult to compare these systems and make progress in the field. Furthermore, based on the reciprocal relationships between neural and symbolic components and how they benefit each other, neuro-symbolic reasoning systems, and in general neuro-symbolic AI systems, can be categorized into one of the five distinct categories [9].

– Symbolic Neuro Symbolic: In this category, input and output are in symbolic form, and processing relies on neural networks, often including Natural Language Processing-based systems.
– Symbolic[Neuro]: Symbolic solvers use neural models internally for some functions, as seen in systems like AlphaGo [10].
– Neuro $\cup$ compile[Symbolic]: These approaches take symbolic rules as input and compile them during training, effectively integrating symbolic knowledge into the structure of neural models, as demonstrated in Deep Learning For Symbolic Mathematics [11].
– Neuro $\rightarrow$ Symbolic: This category involves a refined integration of neural and symbolic approaches, where both systems collaborate to enhance specific tasks, such as in the case of Neuro-Symbolic Concept-Learner [12].
– Neuro[Symbolic]: Refers to the embedding of symbolic reasoning inside a neural engine, such as Graph Neural Networks (GNN).

Each of these categories represents a unique approach to neuro-symbolic AI, adding an extra layer of diversity to the advancements in this field.

Drawing inspiration from Jim Gray's pioneering work [13] on domain-specific benchmarks for databases, our goal is to tackle the challenge of benchmarking neuro-symbolic reasoners. The primary purpose of such a benchmark is two-fold. Firstly, it serves as a tool to identify the performance bottlenecks, enabling targeted improvements in the systems where algorithms are still evolving. Secondly, benchmarks facilitate meaningful comparisons between various systems, offering insights into their relative strengths and weaknesses. While this paper does not put forth an alternative benchmark, we highlight the strong need for such benchmarks, including their features, and explain why they are essential for moving the field forward.

In the following section (Section 2), we delve into the recent advancements in neuro-symbolic reasoning, highlighting the challenges in evaluating and comparing the existing state-of-the-art neuro-symbolic reasoners. Subsequently, in Section 3, we address the barriers that must be overcome to facilitate the effective evaluation of neuro-symbolic reasoners.

2. Neuro-Symbolic Reasoning for Description Logics

In recent years, there have been significant advancements in developing neuro-symbolic reasoners for description logics (DLs) [14], a formal underpinning for the Web Ontology Language (OWL 2) [15]. While most of these works predominantly focus on classification and consistency checking [16–18], the other reasoning tasks, such as instance retrieval, query rewriting, materialization, abduction, and explanation generation, remain relatively unexplored. The intricacy of these tasks varies significantly, and delving into their complexities offers a promising avenue for further exploration.

Research in this domain takes an alternative approach to traditional reasoning tasks such as classification and consistency, breaking them into class subsumption, class membership, and satisfiability tasks. Various techniques are employed, such as geometric embeddings [19–22] that map ontological relationships to geometric spaces and emulating logical reasoning through machine learning [23–25]. A comprehensive overview and detailed insights into the state-of-the-art neuro-symbolic reasoning landscape are discussed in [17, 18]. Regarding other categories, a limited amount of work, such as that for e-commerce search [26], merges neuro-symbolic reasoning with query rewriting. This involves a Knowledge Graph (KG) [27] enhanced neural network approach that integrates auxiliary knowledge from a product Knowledge Graph, enhancing semantic understanding of user queries and improving query reformulation.

The existing traditional benchmarks such as LUBM (Lehigh University Benchmark) [28], UOBM (University Ontology Benchmark) [29], and OWL2Bench [30] lack suitability for evaluating neuro-symbolic reasoners due to their narrow focus on conventional reasoning tasks. Traditional evaluations of reasoning systems often rely on metrics such as reasoning time, which may not align well with the evaluation requirements of neuro-symbolic reasoners. Although the ontologies of these benchmarks, along with those from the OWL Reasoner Evaluation (ORE) Competition [31], can serve as initial datasets for the proposed neuro-symbolic benchmark framework, these datasets fall short of addressing the distinct challenges posed by neuro-symbolic reasoning. To our knowledge, no benchmarks or evaluation frameworks have been designed to evaluate and compare neuro-symbolic reasoning systems. Most reasoner evaluations are performed on different publicly available ontologies, including but not restricted to SNOMED CT[1], Gene Ontology (GO)[2], and Galen[3], as well as other ontologies available in public repositories such as DBpedia [32], YAGO [33], Wikidata [34], Claros[4], NCBO Bioportal[5], and AgroPortal[6]. However, these offer a limited set of ontologies for evaluation, which does not cover the full spectrum of possible scenarios.

As discussed in Section 1, neuro-symbolic approaches encompass a range of evaluation methodologies and reasoning techniques. This diversity becomes evident in Table 1, highlighting the necessity for a dedicated benchmark to systematically and comprehensively assess the performance of neuro-symbolic reasoning systems. The table reveals the utilization of subsets of description logics, such as $\mathcal{ALC}$ and $\mathcal{EL}^{++}$, and various OWL 2 profiles like EL and RL [35]. Some works also incorporate RDF and RDFS into their reasoning techniques, underlining the diversity in the supported ontology languages and profiles, which implies that existing works handle different levels of complexity. Furthermore, the table showcases the variety of reasoning tasks undertaken, different datasets utilized, and the diverse metrics employed for evaluating each approach. The summary column in Table 1 highlights the differences in techniques used by each work. It is important to note that the paper does not aim to provide an exhaustive list of all the existing work. Instead, it emphasizes the variations in reasoning and evaluation approaches. The collective representation highlights the pressing need for a standardized benchmark to facilitate fair and consistent comparisons, thereby advancing the progress of neuro-symbolic reasoning research. The table reveals that similar works may differ significantly by employing distinct metrics and datasets to evaluate their contributions. For instance, consider the works of Makni et al. [25] and Ebrahimi et al. [24]. Both studies focus on RDFS entailment reasoning, aiming to replicate deductive reasoning processes. However, they adopt different metrics and datasets to assess the effectiveness and performance of their approaches. Such variations in evaluation criteria can lead to diverse insights and perspectives on the contributions within the field.

Paper	**Logic**	**Reasoning Task**	**Datasets Used**	**Metrics**	**Summary of Approaches Used**
ELEm [19]	$\mathcal{EL}^{++}$	Subsumption	GO	Hits@n, AUC, Mean Rank	To capture entity relationships, embeddings were created by representing Concepts as n-balls and the relations as translation vectors between the centers of each Concept ball. The embeddings were utilized to predict protein-protein interactions.

[1]https://bioportal.bioontology.org/ontologies/SNOMEDCT
[2]https://bioportal.bioontology.org/ontologies/GO
[3]https://bioportal.bioontology.org/ontologies/GALEN
[4]https://www.clarosnet.org
[5]https://bioportal.bioontology.org/
[6]http://agroportal.lirmm.fr/

EmEL^{++} [20]	$\mathcal{EL}^{++}$	Subsumption	SNOMED CT, Anatomy, GO, Galen	Hits@n, AUC, Median Rank, 90^{th} percentile rank	Extended ELEm with relation inclusion and role chains. Also introduced negative samples for training.
EmEL-V [21]	$\mathcal{EL}^{++}$	Subsumption	SNOMED CT, GO, Galen	Top@n, Median Rank, 90^{th} percentile rank	Extended EmEL^{++} to include many-to-many relationships
BoxEL [22]	$\mathcal{EL}^{++}$	Subsumption	Anatomy, GO, Galen	Hits@n, AUC, Mean Rank	To capture entity relationships, mapped concepts as boxes and deals with the limitations of n-ball [19–21] based embeddings.
Özçep et al. [36]	$\mathcal{ALC}$	Concept Membership	NA	NA	Embeds Concepts in the ontology as convex regions in vector spaces.
E2R [37]	$\mathcal{ALC}$	Concept Membership	LUBM	Hits@n, Mean Rank, MRR	Aiming to preserve the logical structure, proposed embeddings in the quantum space.
Makni and Hendler [25]	RDFS	Entailment Reasoning	LUBM and Scientist dataset created from DBpedia	Precision, recall, and F1-score	The evaluation focused on assessing noise tolerance by employing an encoder-decoder architecture to translate input RDF graph embeddings into corresponding inference graph embeddings.
Ebrahimi et al. [38]	RDFS	Query-based Classification	Created from Linked Data Cloud and Data Hub websites	Precision, recall, and F1-score	Explored the capabilities of end-2-end memory networks. The model's capability for multi-hop reasoning is demonstrated. The use of normalized embeddings support transfer.
Ebrahimi et al. [24]	RDFS and $\mathcal{EL}^{+}$	Entailment Reasoning	Synthetic Data and LUBM	Exact Matching Accuracy	Utilized pointer networks for learning the sequential application of inference rules used in many deductive reasoning algorithms.
Hohenecker and Lukasiewicz [39]	OWL 2 RL	Entailment Reasoning	Claros, DBpedia, UMLS, and Synthetic Data	Accuracy	Developed a deep learning-based model called Recursive Reasoning Networks (RNN).

Eberhart et al. [23]	$\mathcal{EL}^+$	Ontology Completion (concept inclusions and existential restrictions)	Synthetic Data and SNOMED	Precision, recall, and F1-score	Showcases completion reasoning behavior using various LSTM neural networks to learn reasoning patterns, employing three distance measures to assess prediction accuracy.
Makni et al. [40]	RDFS	Explainable Entailment Reasoning	LUBM and real-world scholarly dataset	Accuracy	Built upon the previous work [25] for generating explanations for the derived conclusions by taking the RDF graph and inferred triples as input and the explanations as the target.
Hohenecker and Lukasiewicz [41]	RDF	Concept Membership and Relation Prediction	LUBM, UOBM, Claros, DBpedia	F1 score and Accuracy	Proposed Relational Tensor Network (RTN). Embeddings of the individuals are computed by applying RTNs on the Directed Acyclic Graph representation of the ontology (including the inferences).
Farzana et al. [26]	RDF	Query pruning and complete query rewriting	Created from user search logs from eBay Inc.	Precision, Recall, and F-score, and query accuracy,	Proposes a Knowledge Graph (KG) enhanced approach for query rewriting in e-commerce, leveraging RDF2Vec entity embeddings, entity types, category information, and entity frequency extracted from a product KG.

Table 1: Overview of Variations in Neuro-Symbolic Reasoning and Evaluation Approaches

To further highlight the diversity in the current approaches, we classify the works mentioned in Table 1 into one of the five distinct categories discussed in Section 1. [23–25, 38, 40, 41] take symbolic reasoning rules as input and compile them during training (Neuro $\cup$ compile[Symbolic]), integrating symbolic knowledge into neural models. [19–22, 36, 37] embed symbolic reasoning inside neural engines, representing symbolic information in geometric or vector spaces and employing neural methods for reasoning tasks (Neuro[Symbolic]). Additionally, [26] falls into the category involving a refined integration of neural and symbolic approaches to enhance query rewriting (Neuro $\rightarrow$ Symbolic).

3. Desiderata for Benchmarking Neuro-Symbolic Reasoners

Creating an effective benchmark demands careful consideration of critical principles such as simplicity for accessibility, portability for impartial assessments across various approaches, scalability to accommodate diverse system sizes, and relevance to reflect practical challenges in benchmark scenarios [13]. However, the evaluation of neuro-symbolic reasoners presents its own set of distinctive challenges. Given the field's novelty, state-of-the-art solutions do not approach such challenges systematically. Therefore, we advocate below the issues that should be prioritized in constructing a fair neuro-symbolic reasoning benchmark.

1. **Diverse benchmark scenarios**
   Neuro-symbolic reasoning is a sophisticated AI method aiming to make sense of complex application domains. Thus, to comprehensively evaluate neuro-symbolic reasoners, the benchmark should represent such complexity by including diverse ontologies that vary in size (Abox and TBox), profile, and axiom types. Different neuro-symbolic reasoners work differently. Therefore, the benchmark scenarios should also include specific reasoning tasks as well as generic information needs that can be implemented differently. While the former paves the road to micro benchmarking, the latter will ensure comparability across heterogeneous neuro-symbolic architectures. Finally, it is important for the benchmark to be scalable and realistic. The benchmark scenarios should follow real-world use cases. However, the benchmark should be able to scale beyond the existing requirements to foster technological progress in neuro-symbolic reasoning.
2. **Introducing controlled inconsistencies**
   An intriguing challenge in benchmark design is to enable the benchmark to generate controlled inconsistencies or even paradoxical scenarios in a deterministic manner. While the importance of this feature is recognized, it's noteworthy that existing benchmarks may not yet possess the capacity to introduce generic inconsistencies in a manner congruent with the context of repair. Additionally, ensuring the reproducibility of such inconsistency generation poses significant difficulties. Surprisingly, conventional generative AI models like Large Language Models [42] are not suitable for this purpose. This highlights the unique demands of creating benchmark scenarios that emulate real-world inconsistencies yet remain controlled and reproducible.
3. **Input representation**
   One of the pivotal challenges in neuro-symbolic reasoning is the representation of information in terms of embeddings. Traditional embedding techniques such as TransE [43], although powerful, were not initially designed to retain logical information. This leads to partitioning the knowledge representation into ABox (assertional knowledge) and TBox (terminological knowledge). To address this, the initial step towards creating a well-founded reasoner involves engineering an embedding method that can effectively capture logical relationships and nuances. However, imposing a specific representation within the benchmark may be counterproductive unless its objective is testing specific neurosymbolic properties. The selection of the appropriate input representation plays a crucial role in ensuring that the benchmark's evaluation environment mirrors real-world scenarios, aligning with the amalgamation of neural networks and logical reasoning.
4. **Assessment of the deductive capabilities of existing approaches**
   In the trajectory towards developing a new generation of reasoners that effectively harness the potential of both neural networks and logical reasoning, a foundational requirement involves conducting an equitable assessment of state-of-the-art solutions. The existing approaches have showcased promising minimal deductive capabilities. It is critical to ascertain how these models can push their deductive boundaries. This assessment provides insights into the present capabilities of these approaches and illuminates the trajectory of the field's future development. By exploring the limits of these models, researchers can understand the potential pathways for enhancing and expanding the capabilities of neuro-symbolic reasoning.
5. **Success metrics and key performance indicators**
   The evaluation of neurosymbolic reasoning necessitates the formulation of metrics that accurately gauge the success of these approaches. Traditional reasoning methods have relied on notions such as soundness, completeness, and correctness to validate the deductions. However, these terms don't fit the realm of neuro-symbolic reasoning due to the approximate nature of their outcomes. To address this gap, there is a compelling need to develop novel metrics that integrate logical principles of soundness and completeness with statistical metrics. Innovative metrics like these would provide a detailed and comprehensive assessment of how well neuro-symbolic approaches perform, blending neural networks with logical reasoning. An ideal neuro-symbolic reasoner should support the most expressive logic, be transferable to different domains, generate all and only correct inferences in a single run, and provide explanations for the generated consequences at scale and with high performance. The benchmark should evaluate the extent to which neuro-symbolic reasoners manifest these desired features. This evaluation is crucial for understanding the limitations of current systems and guiding future research directions.

4. Conclusion and Future Work

We highlighted the significant need for a comprehensive benchmark framework to tackle the challenges tied to evaluating neuro-symbolic reasoning systems. Merging symbolic logic and neural network-based machine learning brings great promise, but the lack of common evaluation methods has held back progress in the field. By underlining the importance of creating benchmarks, our aim for the future is to establish a structured way of evaluating these systems that can drive the field forward.

When constructing the benchmark, it is important to select reasoning tasks strategically, focusing on challenges that significantly impact the advancements in the field. Prioritize tasks that are complex and represent the real-world. The benchmark suite should resemble a versatile toolbox, generating diverse challenges tailored to each task. Given the rapid evolution of the neuro-symbolic reasoning domain, the adaptability of the benchmark holds significant importance. It should seamlessly integrate new tasks, ensuring the benchmark remains up-to-date and pertinent.

Acknowledgement

Gunjan Singh and Raghava Mutharaju would like to acknowledge the partial support of the Infosys Centre for Artificial Intelligence (CAI), IIIT-Delhi, in this work.

References

[1] A.S. d'Avila Garcez, T.R. Besold, L.D. Raedt, P. Földiák, P. Hitzler, T. Icard, K. Kühnberger, L.C. Lamb, R. Miikkulainen and D.L. Silver, Neural-Symbolic Learning and Reasoning: Contributions and Challenges, in: *2015 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 22-25, 2015*, AAAI Press, 2015. http://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10281.

[2] M.K. Sarker, L. Zhou, A. Eberhart and P. Hitzler, Neuro-Symbolic Artificial Intelligence, *AI Communications* **34**(3) (2021), 197–209.

[3] J. Ott, A. Ledaguenel, C. Hudelot and M. Hartwig, How to Think About Benchmarking Neurosymbolic AI?, in: *Proceedings of the 17th International Workshop on Neural-Symbolic Learning and Reasoning, La Certosa di Pontignano, Siena, Italy, July 3-5, 2023*, A.S. d'Avila Garcez, T.R. Besold, M. Gori and E. Jiménez-Ruiz, eds, CEUR Workshop Proceedings, Vol. 3432, CEUR-WS.org, 2023, pp. 248–254. https://ceur-ws.org/Vol-3432/paper22.pdf.

[4] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li and M. Sun, Graph neural networks: A review of methods and applications, *AI Open* **1** (2020), 57–81. doi:10.1016/j.aiopen.2021.01.001.

[5] S. Badreddine, A.S. d'Avila Garcez, L. Serafini and M. Spranger, Logic Tensor Networks, *Artif. Intell.* **303** (2022), 103649. doi:10.1016/j.artint.2021.103649.

[6] P. Sen, B.W.S.R. de Carvalho, R. Riegel and A.G. Gray, Neuro-Symbolic Inductive Logic Programming with Logical Neural Networks, in: *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, AAAI Press, 2022, pp. 8212–8219. https://ojs.aaai.org/index.php/AAAI/article/view/20795.

[7] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao and J. Zhu, Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges, in: *Natural Language Processing and Chinese Computing - 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9-14, 2019, Proceedings, Part II*, J. Tang, M. Kan, D. Zhao, S. Li and H. Zan, eds, Lecture Notes in Computer Science, Vol. 11839, Springer, 2019, pp. 563–574. doi:10.1007/978-3-030-32236-6_51.

[8] J. Chen, F. Lécué, Y. Geng, J.Z. Pan and H. Chen, Ontology-guided Semantic Composition for Zero-shot Learning, in: *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR 2020, Rhodes, Greece, September 12-18, 2020*, D. Calvanese, E. Erdem and M. Thielscher, eds, 2020, pp. 850–854. doi:10.24963/kr.2020/87.

[9] H.A. Kautz, The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture, *AI Magazine* **43**(1) (2022), 93–104. doi:10.1609/aimag.v43i1.19122.

[10] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T.P. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel and D. Hassabis, Mastering the game of Go with deep neural networks and tree search, *Nat.* **529**(7587) (2016), 484–489. doi:10.1038/nature16961.

[11] G. Lample and F. Charton, Deep Learning For Symbolic Mathematics, in: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, OpenReview.net, 2020. https://openreview.net/forum?id=S1eZYeHFDS.

[12] J. Mao, C. Gan, P. Kohli, J.B. Tenenbaum and J. Wu, The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision, in: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, OpenReview.net, 2019. https://openreview.net/forum?id=rJgMlhRctm.

[13] J. Gray (ed.), *The Benchmark Handbook for Database and Transaction Systems (2nd Edition)*, Morgan Kaufmann, 1993. ISBN 1-55860-292-5.

[14] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi and P.F. Patel-Schneider (eds), *The Description Logic Handbook: Theory, Implementation, and Applications*, Cambridge University Press, 2003. ISBN 0-521-78176-0.

[15] B.C. Grau, I. Horrocks, B. Motik, B. Parsia, P.F. Patel-Schneider and U. Sattler, OWL 2: The next step for OWL, *Journal of Web Semantics* **6**(4) (2008), 309–322. doi:10.1016/j.websem.2008.05.001.

[16] P. Hitzler, M. Krötzsch and S. Rudolph, *Foundations of Semantic Web Technologies*, Chapman and Hall/CRC Press, 2010. ISBN 9781420090505. http://www.semantic-web-book.org/.

[17] B. Makni, M. Ebrahimi, D. Gromann and A. Eberhart, Neuro-Symbolic Semantic Reasoning, in: *Neuro-Symbolic Artificial Intelligence: The State of the Art*, P. Hitzler and M.K. Sarker, eds, Frontiers in Artificial Intelligence and Applications, Vol. 342, IOS Press, 2021, pp. 253–279. ISBN 978-1-64368-244-0. doi:10.3233/FAIA210358.

[18] G. Singh, S. Bhatia and R. Mutharaju, Neuro-Symbolic RDF and Description Logic Reasoners: The State-Of-The-Art and Challenges, in: *Compendium of Neurosymbolic Artificial Intelligence*, P. Hitzler and M.K. Sarker, eds, Frontiers in Artificial Intelligence and Applications, Vol. 369, IOS Press, 2023, pp. 29–63. ISBN 978-1-64368-407-9. doi:10.3233/FAIA230134.

[19] M. Kulmanov, W. Liu-Wei, Y. Yan and R. Hoehndorf, EL Embeddings: Geometric Construction of Models for the Description Logic $\mathcal{EL}^{++}$, in: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, ed., ijcai.org, 2019, pp. 6103–6109. doi:10.24963/ijcai.2019/845.

[20] S. Mondal, S. Bhatia and R. Mutharaju, EmEL^{++}: Embeddings for $\mathcal{EL}^{++}$ Description Logic, in: *Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021), Stanford University, Palo Alto, California, USA, March 22-24, 2021*, A. Martin, K. Hinkelmann, H. Fill, A. Gerber, D. Lenat, R. Stolle and F. van Harmelen, eds, CEUR Workshop Proceedings, Vol. 2846, CEUR-WS.org, 2021. http://ceur-ws.org/Vol-2846/paper19.pdf.

[21] B. Mohapatra, S. Bhatia, R. Mutharaju and G. Srinivasaraghavan, Why Settle for Just One? Extending $\mathcal{EL}^{++}$ Ontology Embeddings with Many-to-Many Relationships, *CoRR* **abs/2110.10555** (2021). https://arxiv.org/abs/2110.10555.

[22] B. Xiong, N. Potyka, T. Tran, M. Nayyeri and S. Staab, Faithful Embeddings for $\mathcal{EL}^{++}$ Knowledge Bases, in: *The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings*, U. Sattler, A. Hogan, C.M. Keet, V. Presutti, J.P.A. Almeida, H. Takeda, P. Monnin, G. Pirrò and C. d'Amato, eds, Lecture Notes in Computer Science, Vol. 13489, Springer, 2022, pp. 22–38. doi:10.1007/978-3-031-19433-7_2.

[23] A. Eberhart, M. Ebrahimi, L. Zhou, C. Shimizu and P. Hitzler, Completion Reasoning Emulation for the Description Logic $\mathcal{EL}^{+}$, in: *Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice, AAAI-MAKE 2020, Palo Alto, CA, USA, March 23-25, 2020, Volume I*, A. Martin, K. Hinkelmann, H. Fill, A. Gerber, D. Lenat, R. Stolle and F. van Harmelen, eds, CEUR Workshop Proceedings, Vol. 2600, CEUR-WS.org, 2020. http://ceur-ws.org/Vol-2600/paper5.pdf.

[24] M. Ebrahimi, A. Eberhart and P. Hitzler, On the Capabilities of Pointer Networks for Deep Deductive Reasoning, *CoRR* **abs/2106.09225** (2021). https://arxiv.org/abs/2106.09225.

[25] B. Makni and J.A. Hendler, Deep learning for noise-tolerant RDFS reasoning, *Semantic Web* **10**(5) (2019), 823–862. doi:10.3233/SW-190363.

[26] S. Farzana, Q. Zhou and P. Ristoski, Knowledge Graph-Enhanced Neural Query Rewriting, in: *Companion Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, ACM, 2023, pp. 911–919. doi:10.1145/3543873.3587678.

[27] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, S. Kirrane, J.E.L. Gayo, R. Navigli, S. Neumaier, A.N. Ngomo, A. Polleres, S.M. Rashid, A. Rula, L. Schmelzeisen, J.F. Sequeda, S. Staab and A. Zimmermann, Knowledge Graphs, *ACM Computing Surveys* **54**(4) (2022), 71:1–71:37. doi:10.1145/3447772.

[28] Y. Guo, Z. Pan and J. Heflin, LUBM: A Benchmark for OWL Knowledge Base Systems, *Journal of Web Semantics.* 3(2–3) (2005), 158–182.

[29] L. Ma, Y. Yang, G. Qiu Z .and Xie, Y. Pan and S. Liu, Towards a Complete OWL Ontology Benchmark, in: *The Semantic Web: Research and Applications*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 125–139.

[30] G. Singh, S. Bhatia and R. Mutharaju, OWL2Bench: A Benchmark for OWL 2 Reasoners, in: *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II*, J.Z. Pan, V.A.M. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne and L. Kagal, eds, Lecture Notes in Computer Science, Vol. 12507, Springer, 2020, pp. 81–96. doi:10.1007/978-3-030-62466-8_6.

[31] B. Parsia, N. Matentzoglu, R.S. Gonçalves, B. Glimm and A. Steigmiller, The OWL Reasoner Evaluation (ORE) 2015 Competition Report, *Journal of Automated Reasoning* **59**(4) (2017), 455–482. doi:10.1007/s10817-017-9406-8.

[32] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer and C. Bizer, DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia, *Semantic Web Journal* **6** (2014). doi:10.3233/SW-140134.

[33] F.M. Suchanek, G. Kasneci and G. Weikum, YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia, in: *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, Association for Computing Machinery, New York, NY, USA, 2007, pp. 697–706–. ISBN 9781595936547. doi:10.1145/1242572.1242667.

[34] D. Vrandečić, Wikidata: A New Platform for Collaborative Data Collection, in: *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, Association for Computing Machinery, New York, NY, USA, 2012, pp. 1063–1064–. ISBN 9781450312301. doi:10.1145/2187980.2188242.

[35] B. Motik, B.C. Grau, I. Horrocks, Z. Wu, A. Fokoue and C. Lutz, OWL 2 Web Ontology Language Profiles (Second Edition), 2012. https://www.w3.org/TR/owl2-profiles/.

[36] Ö.L. Özçep, M. Leemhuis and D. Wolter, Cone Semantics for Logics with Negation, in: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, C. Bessiere, ed., ijcai.org, 2020, pp. 1820–1826. doi:10.24963/ijcai.2020/252.

[37] D. Garg, S. Ikbal, S.K. Srivastava, H. Vishwakarma, H.P. Karanam and L.V. Subramaniam, Quantum Embedding of Knowledge for Reasoning, in: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, H.M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E.B. Fox and R. Garnett, eds, 2019, pp. 5595–5605. https://proceedings.neurips.cc/paper/2019/hash/cb12d7f933e7d102c52231bf62b8a678-Abstract.html.

[38] M. Ebrahimi, M.K. Sarker, F. Bianchi, N. Xie, D. Doran and P. Hitzler, Reasoning over RDF Knowledge Bases using Deep Learning, *CoRR* **abs/1811.04132** (2018). http://arxiv.org/abs/1811.04132.

[39] P. Hohenecker and T. Lukasiewicz, Ontology Reasoning with Deep Neural Networks, *Journal of Artificial Intelligence Research* **68** (2020), 503–540. doi:10.1613/jair.1.11661.

[40] B. Makni, I. Abdelaziz and J.A. Hendler, Explainable Deep RDFS Reasoner, *CoRR* **abs/2002.03514** (2020). https://arxiv.org/abs/2002.03514.

[41] P. Hohenecker and T. Lukasiewicz, Deep Learning for Ontology Reasoning, *CoRR* **abs/1705.10342** (2017). http://arxiv.org/abs/1705.10342.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser and I. Polosukhin, Attention is All you Need, in: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, H.M. Wallach, R. Fergus, S.V.N. Vishwanathan and R. Garnett, eds, 2017, pp. 5998–6008. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[43] A. Bordes, N. Usunier, A. García-Durán, J. Weston and O. Yakhnenko, Translating Embeddings for Modeling Multi-relational Data, in: *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, C.J.C. Burges, L. Bottou, Z. Ghahramani and K.Q. Weinberger, eds, 2013, pp. 2787–2795. https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html.